General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

RR-Y1-04217
NSTL/ERL-202

AgRISTARS

E82-10287
TM-84696

A Joint Program for
Agriculture and
Resources Inventory
Surveys Through
Aerospace
Remote Sensing

Renewable Resources Inventory

March 1982

Technical Report

Original photography may be purchased
from EROS Data Center
Sioux Falls, SD 57198

ANALYSIS OF THEMATIC MAPPER SIMULATOR DATA ACQUIRED DURING WINTER SEASON OVER PEARL RIVER, MS, TEST SITE

J. E. Anderson
M. T. Kalcic

National Aeronautics and Space Administration
National Space Technology Laboratories
Earth Resources Laboratory
NSTL Station, MS 39529

(E82-10287) ANALYSIS OF THEMATIC MAPPER SIMULATOR DATA ACQUIRED DURING WINTER SEASON OVER PEARL RIVER, MISSISSIPPI, TEST SITE (NASA) 45 p HC A03/MF A01 CSCL 05B
N82-24565
Unclas
G3/43 00287

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
RR-Y1-04217 NSTL/ERL-202		

4. TITLE AND SUBTITLE	5. REPORT DATE
Analysis of Thematic Mapper Simulator Data Acquired During Winter Season over Pearl River, MS Test Site	March 1982
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Anderson, J.E. and M.T. Kalcic	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NO.
NASA/National Space Technology Laboratories Earth Resources Laboratory NSTL Station, MS 39529	11. CONTRACT OR GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT & PERIOD COVERED
National Aeronautics and Space Administration	Technical Report
	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

This report presents results of digital processing of aircraft-acquired Thematic Mapper Simulator (TMS) data collected during the winter season over a forested site in southern Mississippi. The goal of the research was to investigate the utility of TMS data for use in forest inventories and monitoring. This study deals with one of four test sites selected for this research task under the AgRISTARS Renewable Resources Inventory Project.

Analyses indicate that TMS data are capable of delineating the mixed forest land cover type to an accuracy of 92.5 percent correct. The accuracies associated with river bottom forest and pine forest were 95.5 and 91.5 percent correct, respectively. These figures reflect the performance for products produced using the best subset of channels for each forest cover type. It was found that the choice of channels (subsets) had a significant effect on the accuracy of classifications produced, and that the same channels are not the most desirable for all three forest types studied. Both supervised and unsupervised spectral signature development techniques were evaluated; the unsupervised methods proved unacceptable for the three forest types considered.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Remote Sensing Thematic Mapper Simulator Landsat-D Pattern Recognition	Unlimited

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
Unclassified	Unclassified	44	*

NSTL FORM 13 (JAN 1975)

*For sale by NTIS, Springfield, VA 22151

ANALYSIS OF THEMATIC MAPPER SIMULATOR DATA ACQUIRED DURING WINTER SEASON OVER PEARL RIVER, MS TEST SITE

MARCH 1982

J.E. ANDERSON

and

M.T. KALCIC

RR-Y1-04217

NSTL/ERL-202

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

NATIONAL SPACE TECHNOLOGY LABORATORIES

EARTH RESOURCES LABORATORY

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

SUMMARY

This report presents results of digital processing of aircraft-acquired Thematic Mapper Simulator (TMS) data collected during the winter season over a forested site in southern Mississippi. The goal of the research was to investigate the utility of TMS data for use in forest inventories and monitoring. This study deals with one of four test sites selected for this research task under the AgRISTARS Renewable Resources Inventory Project.

Analyses indicate that TMS data are capable of delineating the mixed forest land cover type to an accuracy of 92.5 percent correct. The accuracies associated with river bottom forest and pine forest were 95.5 and 91.5 percent correct, respectively. These figures reflect the performance for products produced using the best subset of channels for each forest cover type. It was found that the choice of channels (subsets) had a significant effect on the accuracy of classifications produced, and that the same channels are not the most desirable for all three forest types studied. Both supervised and unsupervised spectral signature development techniques were evaluated; the unsupervised methods proved unacceptable for the three forest types considered.

INTRODUCTION

This study was conducted as part of a research task under the AgRISTARS Renewable Resources Inventory (RRI) Project. The overall objectives of the research task are (1) to design and implement an efficient procedure for processing and analysis of Thematic Mapper (TM) digital data, and (2) to examine the potential utility of TM digital data in forest inventories and monitoring. To adequately evaluate the utility of TM digital data for forest management or inventorying/monitoring, analysis should not be restricted to any one major

forest ecosystem, but should include numerous forest types of major importance to a user or user community. All results obtained would thus be indicative of the general nature of the forest resource, with specific problems or considerations related to each major forest type dealt with on a site-by-site level.

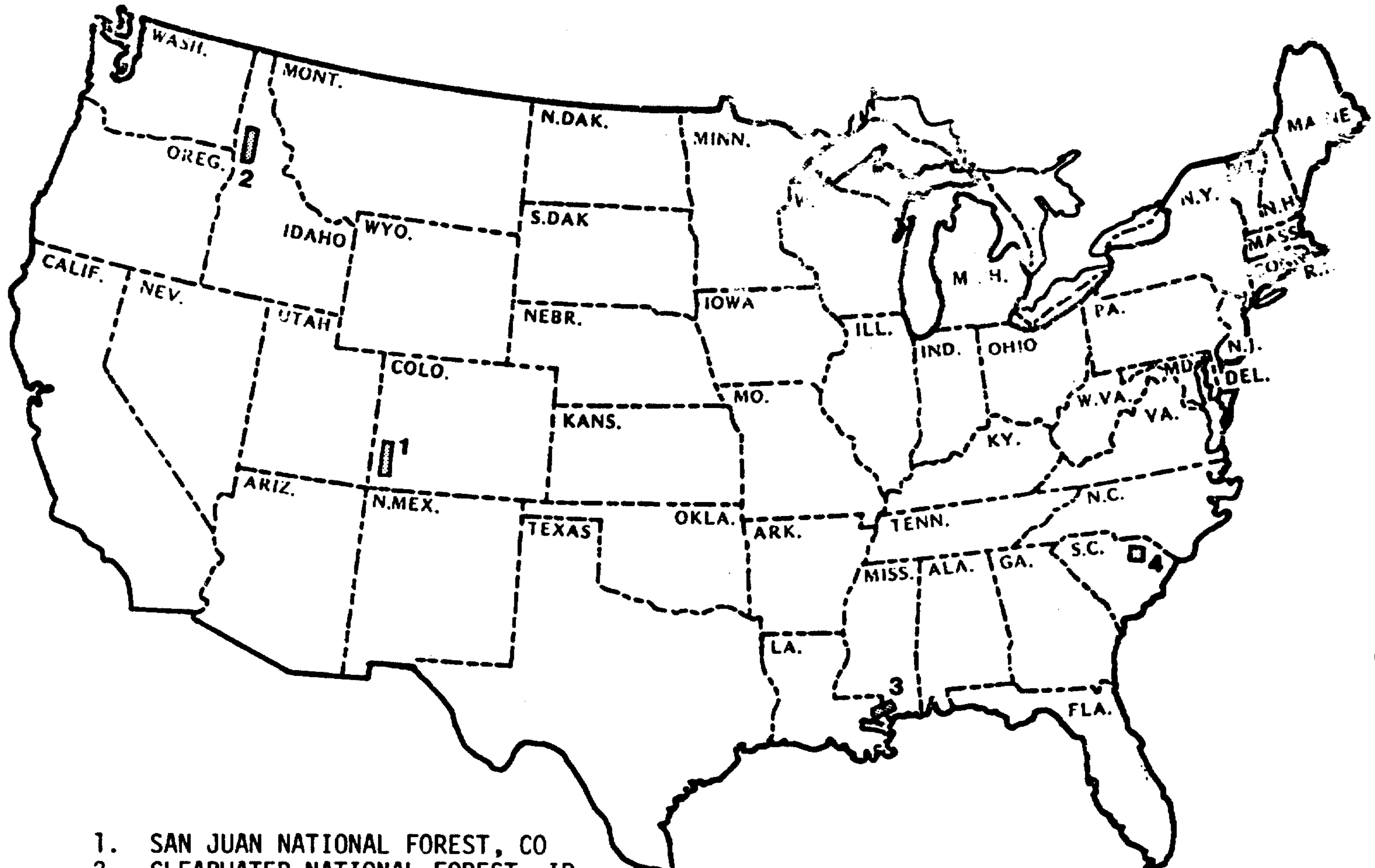
Similarly, since the forest is a dynamic ecosystem with an annual biotic cycle, the effects of phenologic season on the results obtained from digital data analysis must be factored into the investigation.

In order to take into account these considerations of variation in forest ecosystems and their response to the different seasons, four study sites were selected in locations indicated in Figure 1. These sites represent distinctly different forest ecosystems, each with unique environmental conditions and forest cover types. In addition, data collection was scheduled to occur at each site during each of the four major phenologic seasons (winter dormancy, spring leaf-out, summer growth, and fall leaf abscission). The calendar dates associated with each phenologic season were defined for each site independently, based on knowledge of the area. The dates selected included only that time frame during which the phenologic condition of the forest ecosystem remained somewhat stable and typified conditions representative for each season.

Until Landsat D is launched (the last quarter of FY82 or first quarter of FY83) and the TM is available, this research task will employ data collected with an aircraft-borne TMS.

This report deals specifically with results obtained from the analysis of winter TMS digital data collected over the Pearl River Basin study site in southern Mississippi. Subsequent reports will present results obtained for data collected for other seasons or for the other three study sites being used in this research. In addition, a comparison of results obtained from the

ORIGINAL PAGE IS
OF POOR QUALITY



1. SAN JUAN NATIONAL FOREST, CO
2. CLEARWATER NATIONAL FOREST, ID
3. PEARL RIVER BASIN, LA/MS
4. KERSHAW CO., SC

Figure 1. Study Sites

Landsat multispectral scanner (MSS) data and TMS data is in progress and will be included in a subsequent report.

PEARL RIVER BASIN STUDY SITE

The Pearl River Basin Study Site (hereafter referred to as the MS site) is a 43.2-km (27-mi) north-south area located in southern Mississippi (Figure 1), and includes portions of Hancock County, MS, and Saint Tammany Parish, LA. This particular area was selected to represent the longleaf-slash pine and oak-gum-cypress forest types (reference 1) which occur throughout much of the South (Figure 2). The longleaf-slash pine type is typified by the occurrence of longleaf pine (Pinus Palustis Mill.) and slash pine (Pinus Elliottii Engelm.), but also includes the other southern pines, oak, and gum. Species commonly found in the oak-gum-cypress type include sweetgum (Liquidambar Styraciflua L.), Laurel oak (Quercus Laurifolia Michx.), American Hornbeam (Carpinus Caroliniana Walt.), American Holly (Ilex Opaca Ait.), Water oak (Q. Nigra L.) and Sweet Bay (Magnolia Virginiana L.) on the "drier" sites, and water tupelo (Nyssa Aquatica Marsh.), bald cypress (Taxodium Distichum Rich.), and numerous species of ash (Fraxinus Spp.) on the "wetter" sites (often having standing water for long periods--reference 2).

The topography of the MS site is flat to gently rolling, with elevations ranging between 5 m (15 ft) in the south and 61 m (200 ft) in the north. The distribution of generalized land cover types within the MS site is presented in Figure 3. The lower elevations encompass the major drainage basins of the Pearl River system in the south and the Hobolochitto River (with numerous branches) in the north (Figure 4). Most of the remaining areas, with the exception of one small city, are occupied by pine forests or have been cleared of all native vegetation and are currently supporting agricultural crops.

ORIGINAL PAGE IS
OF POOR QUALITY

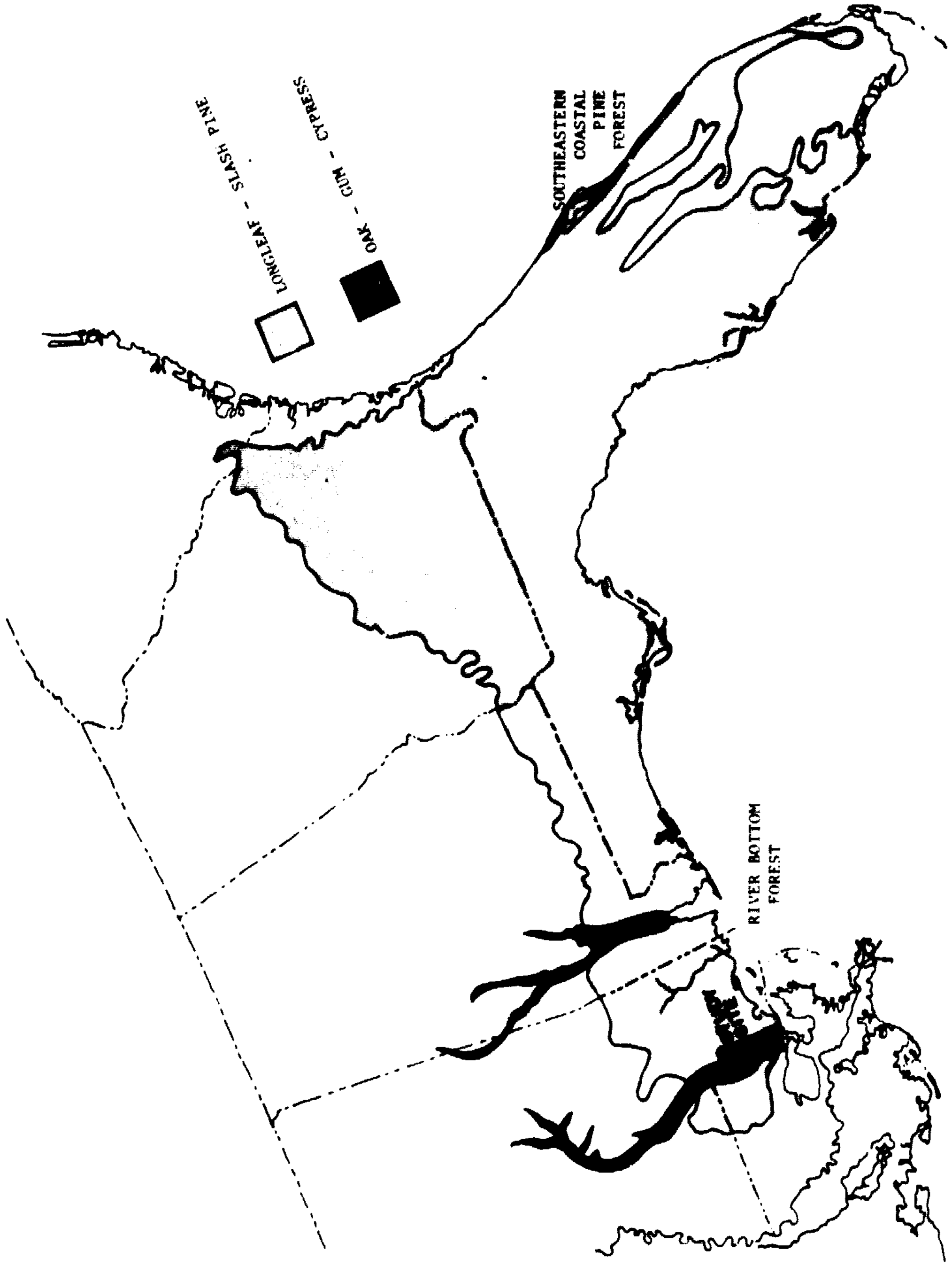


Figure 2. Typical Southeastern U.S. Forests

ORIGINAL PAGE IS
OF POOR QUALITY

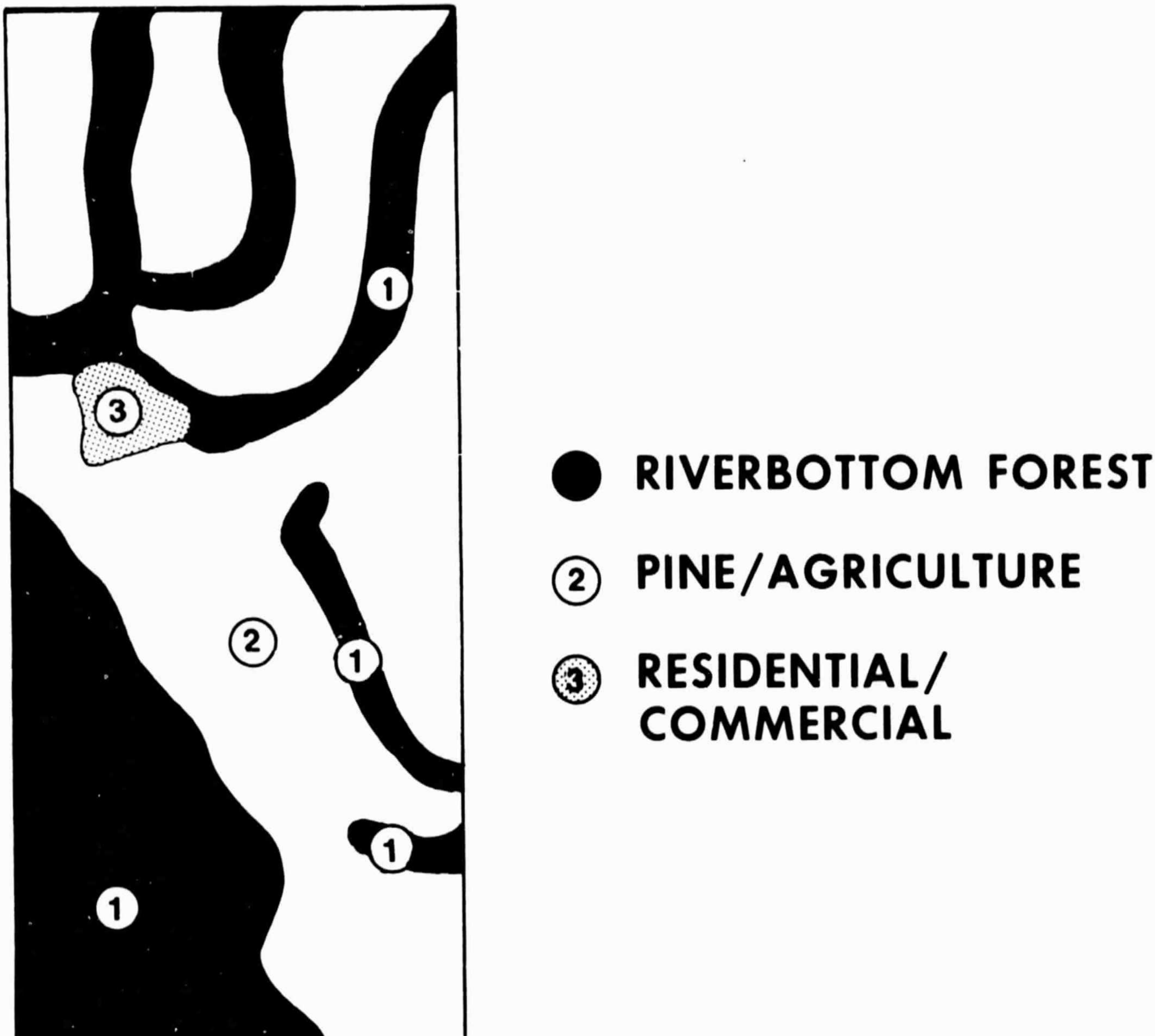


Figure 3. MS Site Land Cover Types

ORIGINAL PAGE IS
OF POOR QUALITY

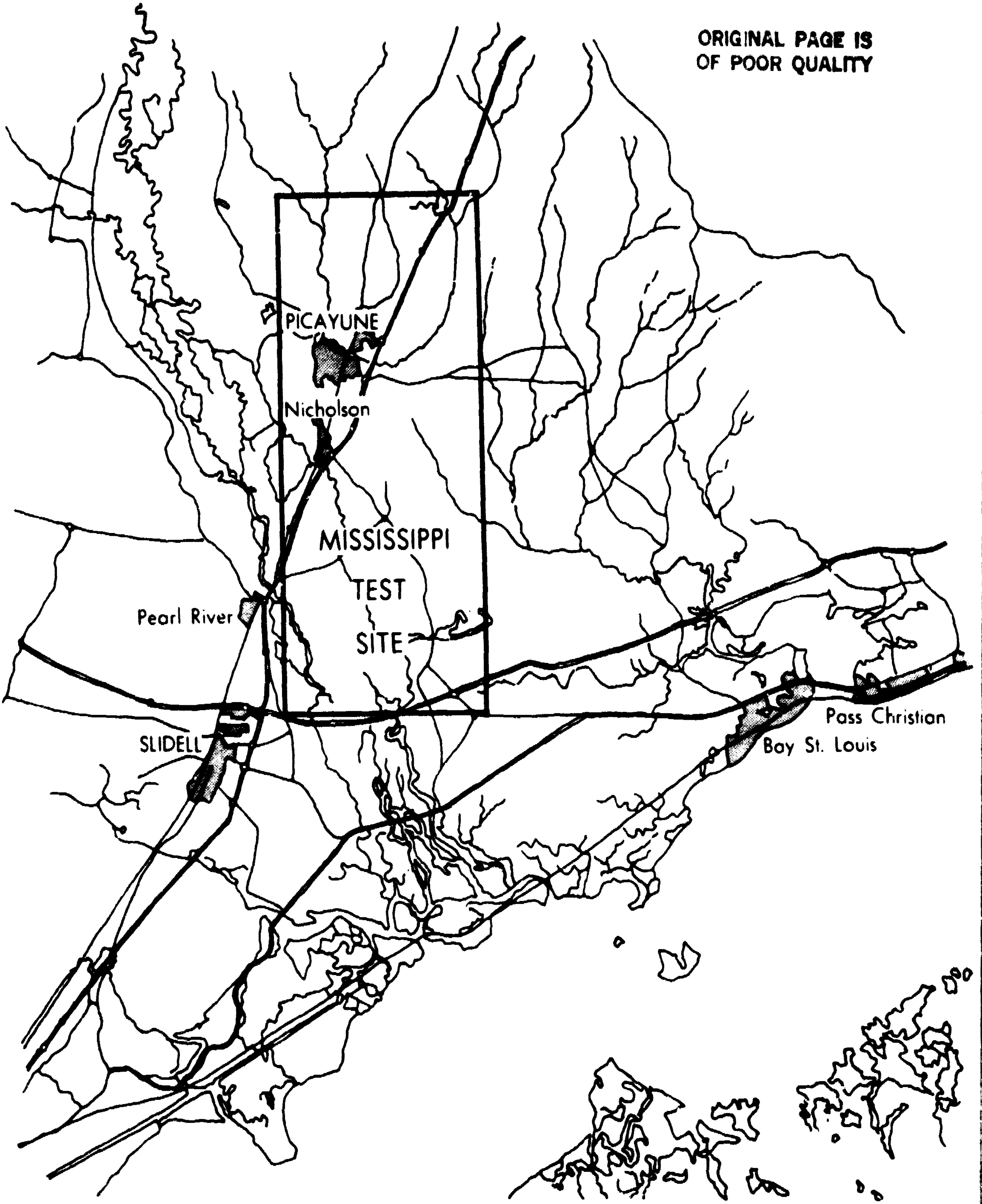


Figure 4. MS Test Site and Surrounding Area

The overall condition of the MS site during the winter season can be ascertained from a careful study of Figure 5. This color infrared aerial photograph covers an area which is typical of the entire MS site and hence will be used as a reference for illustrating statements related to land cover conditions.

All of the land in the MS site can be grouped into eight basic types. A brief discussion dealing with each type is presented below to familiarize the reader with winter land cover conditions (letters in parentheses refer to areas marked in Figure 5):

1. Inert Materials - This general land cover type contains such cover types as sand bars (D), bare soil/gravel pits (E), highways and dirt roads (F), and fallow agricultural fields. Parking lots, cities, buildings, etc., are also included in this land cover type. The physical condition of this land cover is, for the most part, independent of season, as little or no vegetative cover exists on most of these areas. However, rain does affect the condition of bare soil, causing it to become darker in color, and fallow agricultural fields will, in the other seasons, support crops.

2. Winter Crops/Pasture - The predominant agricultural land use during the winter season in the MS site is pasture (H) or winter rye grass (I). Most other agricultural areas are fallow and have been included in the inert materials land cover type. Winter rye grass is very lush and green and, as can be seen on Figure 5, is represented by a bright pink/red on color infrared photography. The pastures are in their characteristic dormant state, and appear brown to the naked eye.

3. Old Fields - This land cover type includes areas which have been cut over and burned (J), or which represent the initial stages of the revegetation

ORIGINAL PAGE
COLOR PHOTOGRAPH

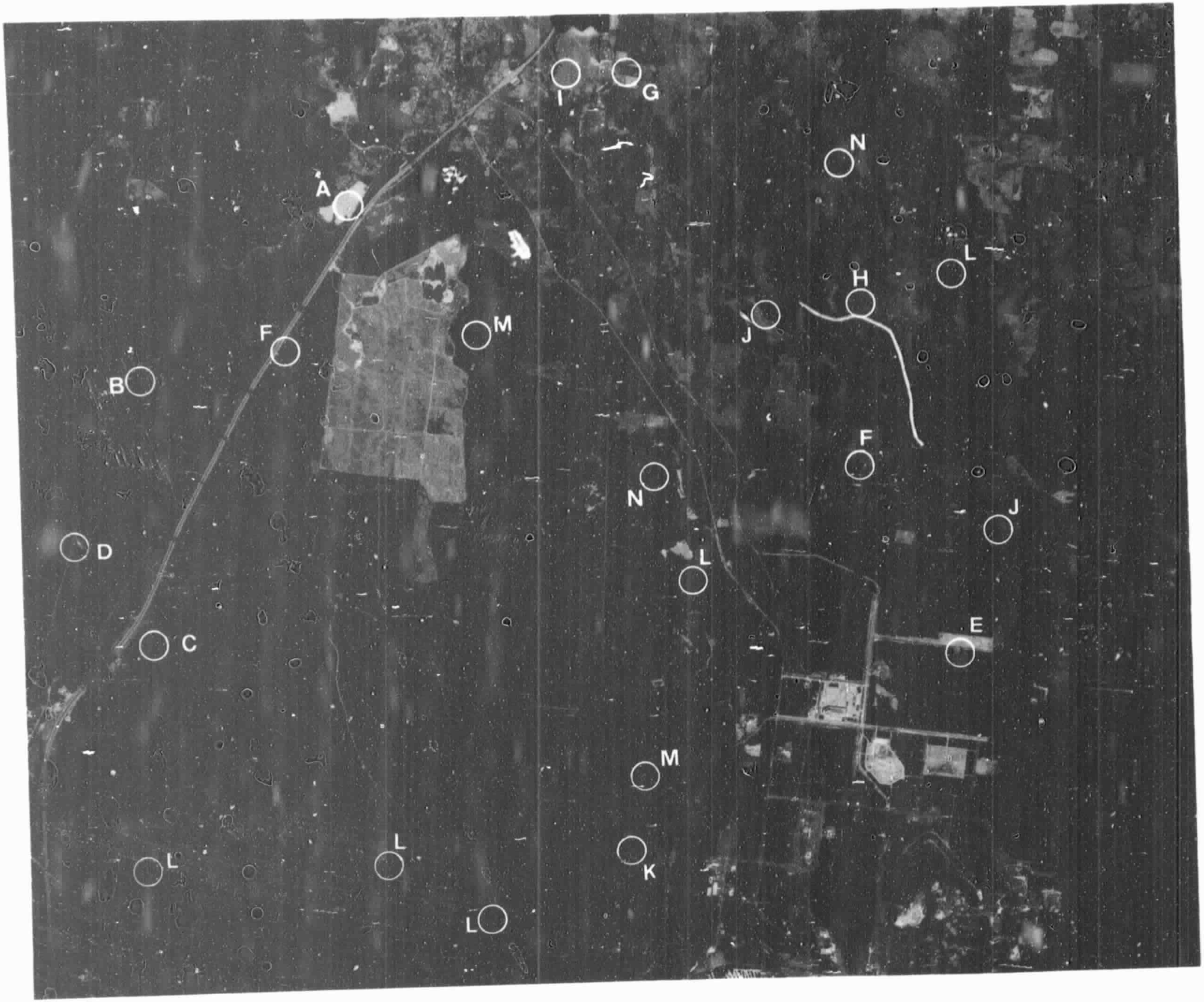


Figure 5. Color IR Photograph of Area Typical of MS Test Site

of abandoned agricultural fields and other cleared areas. Depending on the plant species which are invading a particular site, the condition of these sites may be quite variable.

4. Marsh - Included in this land cover type are areas which are inundated by standing water for a majority of the year, and contain plant species typical of non-forested wetlands covering at least 10 percent of the surface area when viewed from above (K). The marsh is dormant during the winter season, and is quite uniform in color when viewed with the naked eye. Water is, in most instances, present as the substrate on which the dead, brown grass is standing.

5. River Bottom Forest - This category includes forested areas (10 percent or more of the surface area covered by tree foliage when in full leaf conditions) that are seasonally flooded for prolonged periods (usually three months or more) or flooded as a result of diurnal tidal action directly or indirectly through water backup; 66-2/3 percent or more of the foliage cover is made up of the foliage of deciduous tree species when viewed from above in the full foliage condition (L). Depending on the understory species composition and the occurrence of "evergreen" broadleafed tree species (e.g., live oak), the condition of the land cover type is somewhat variable. Also, the amount and condition of standing water when viewed through the vegetation will influence the overall spectral condition of this cover type, especially in the dormant, leafless condition of winter.

6. Coniferous Forest - This includes forested areas with at least 66-2/3 percent of their foliage occupied by coniferous tree species when viewed from above (N). This includes both natural (unmanaged) stands as well as plantations (managed) of coniferous tree species not associated with the river bottom land cover type. Bald cypress and spruce pine (Pinus glabra Walt.)

are not included in this land cover type, since they are considered a natural component of river bottom forest.

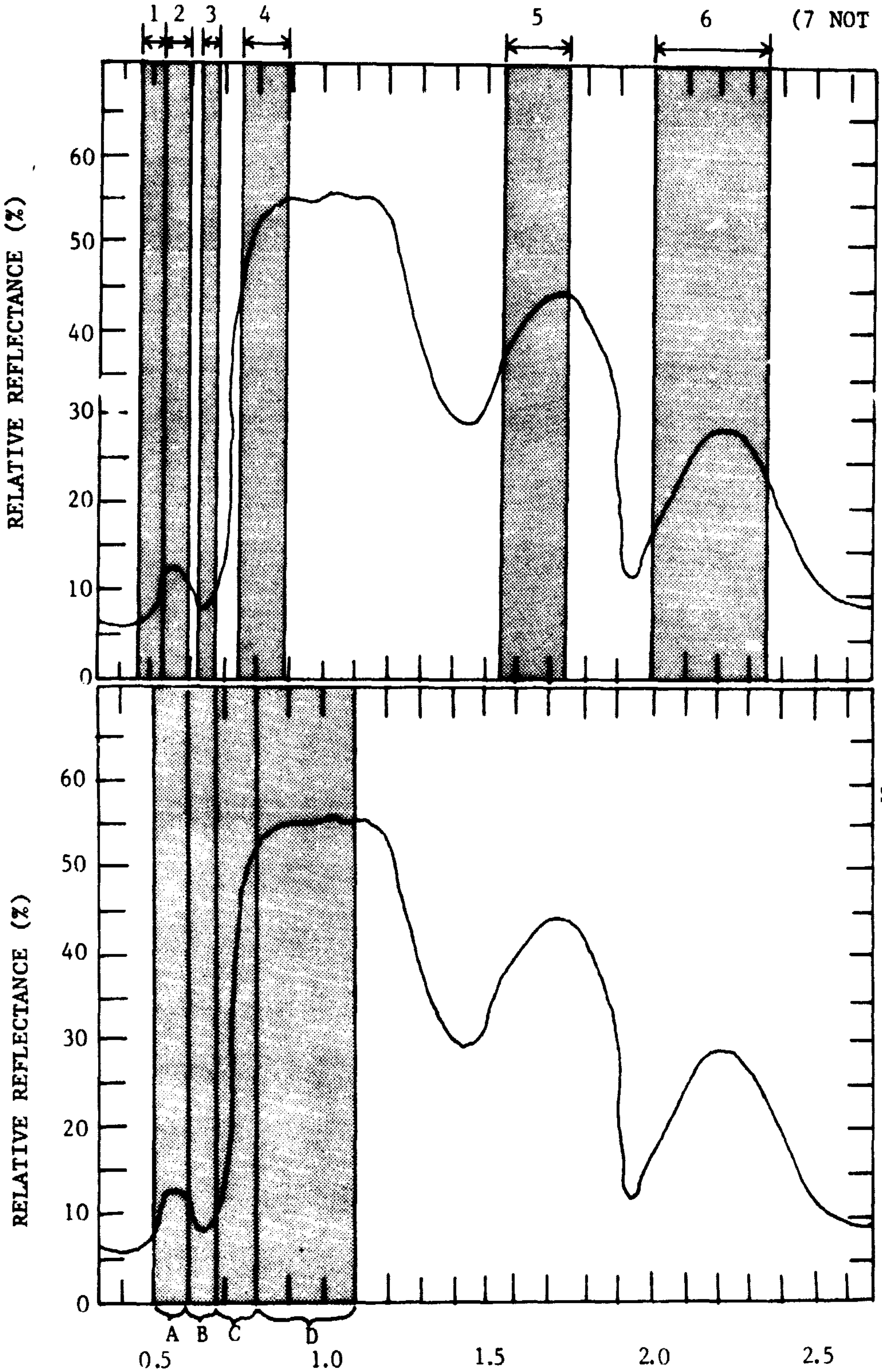
7. Mixed Forest - This land cover type includes forested areas with neither river bottom nor coniferous forest types constituting 66-2/3 percent of the foliage cover when viewed from above in the full foliage condition (M). This land cover type is quite variable in composition, both in terms of overstory species diversity and density of stocking. In addition, where mixed stands do exist, a diversity of understory species which varies by site adds to the overall complexity of the site as seen from above.

8. Water - Most of the water occurring in the MS site exists in one of three conditions, depending to a large degree on the turbidity of the water body in question. "A" in Figure 5 represents water located in a very shallow, sandy borrow pit which, due to the action of wind, waves, and rain, is extremely turbid; "B" represents water which is moderately turbid (the West Pearl River in this example); and "C" depicts very clear water, either shallow or deep. It is not always obvious as to which condition should be assigned to a particular body of water, however. As can be seen in Figure 5, a majority of water is in small, widely scattered ponds, as well as narrow, meandering streams and rivers. In addition, water is present in the understory of much of the river bottom forest area.

THEMATIC MAPPER SIMULATOR DATA

Data used in this study were obtained by an airborne TMS. The TMS was designed to produce data with spectral and spatial characteristics identical to those of the Thematic Mapper on Landsat-D, scheduled for launch later in FY82. This sensor will have spectral resolution as shown in Figure 6, with

ORIGINAL PAGE IS
OF POOR QUALITY



THEMATIC MAPPER SIMULATOR
SPECTRAL RESOLUTION

BAND	WAVELENGTH REGION
1.	.45 - .52μ
2.	.52 - .60μ
3.	.63 - .69μ
4.	.76 - .90μ
5.	1.55 - 1.75μ
6.	2.08 - 2.35μ
7.	10.40 - 12.5 μ

LANDSAT MSS
SPECTRAL RESOLUTION

BAND	WAVELENGTH REGION
A.	.50 - .60μ
B.	.60 - .70μ
C.	.70 - .80μ
D.	.80 - 1.1 μ

Figure 6. Spectral Wavelength Characteristics of TMS and Landsat MSS Systems

30-m (100-ft) spatial resolution in all channels except channel 7, which will have 120-m (394-ft) resolution. Figure 6 also presents the spectral resolution of currently available Landsat MSS data, as well as a generalized green leaf reflectance curve for comparison of the two satellite systems.

Data collected by the TMS are subsequently converted from the analog format produced by the scanner to an 8-bit (256 levels of grey) digital format for use in data processing and analysis activities. The TMS, with 2.5-milliradian aperture, is configured in such a way as to encompass a 50-degree field of view on either side of nadir. Assuming optimal conditions of a perfectly flat target surface and no scan line overlap at nadir, the dimensions of a pixel (at 50 degrees of scan from nadir) would be:

width (across track) = 73.96 m (243 ft)

length (nadir side) = 46.87 m (154 ft)

length (extreme side) = 46.96 m (154 ft)

These dimensions are calculated for an aircraft altitude which would result in a 30 x 30-m (100 x 100-ft) pixel at nadir, the normal Instantaneous Field of View (IFOV) for the TM satellite system. In addition, the overall length of atmosphere through which electromagnetic energy would have to travel across track in order to be measured by the detector is, at 50 degrees, 1.56 times that at nadir. It should be obvious that these conditions are extreme enough to preclude processing and analysis at such large angles of look. These geometric considerations are not as critical at smaller angles of look, and at or near 30 degrees on either side of nadir, the dimensions of the pixels become acceptably close to 30 m (100 ft). Therefore, data processing and analysis were restricted to 30 degrees on either side of nadir, which resulted in a data set containing 1184 scan lines and 418 elements (209 elements on either side of nadir).

TMS data were collected on February 11, 1981, from an altitude of 12,000 m (39,370 ft) above mean terrain elevation. As noted previously, this resulted in a spatial resolution (at nadir) of 30 x 30 m (100 x 100 ft) for channels 1 through 6, and 120 x 120 m (394 x 394 ft) for channel 7. Note that TMS channels are numbered in the sequence that they occur in the electromagnetic spectrum, and are not the same number assignment as for the TM. The data were viewed on an image display device, and examined for radiometric fidelity and the presence of abnormal data values (detector noise, drop outs, loss of synchronization, etc.). It was determined at this time that TMS channel 6 contained an unacceptably high amount of streaking in the data, traced to problems with the detector used in the scanner. Such problems appeared on a black and white display of the data as "comb" marks running across the data scan lines. The coefficient of variation for channel 6 was not found to be markedly different from the other channels (Table 1), and an evaluation of a histogram for channel 6 showed no aberrant behavior Therefore, due to the amount of noise present, channel 6 was removed from this study.

The spatial resolution of channel 7, being four times that of the remaining six channels, resulted in the seventh channel containing only one-sixteenth the number of pixels as the other six channels. This situation was rectified by expanding the data for channel 7 in order to fill up the entire file. Thus, each "pixel" in channel 7 was repeated three times in both the scan line and element directions, resulting in a block of data (four by four pixels in size) containing one radiometric value for all 16 pixels. The value assigned to the pixels was that of the initial channel 7 pixel which was expanded. In this manner, a channel-to-channel registra-

tion with the other six channels was effected, while at the same time preserving the geometric relationship of channel 7 to the other channels (four to one).

When all problems had been corrected, the center 60 degrees of the data were examined for sun angle/angle of look related trends. No such problems existed with this data set, as the aircraft data collection flight (oriented north/south) occurred within one half hour of local solar noon. Figures 7 through 12 show a representative portion of the test site with images produced from TMS data from each of the six channels of TMS data used for this study.

GROUND TRUTH

In order to establish the level of performance of TMS digital data, a network of 136 ground truth sites was established, against which results of the analysis of TMS data could be compared. The ground truth sites were selected to represent the major land cover types of interest within the MS site. Numerous ground truth sites were taken for each land cover type of interest to insure the statistical reliability of results obtained.

Each ground truth site was visited in the field, and a detailed survey was made of the land cover type present. Using photographs taken by field personnel, as well as the written descriptions completed for each site visited, all sites were assigned to one of the eight basic land cover types. Sites retained for use after being visited in the field were transferred onto a small scale vertical photograph. Ground truth corresponding to these sites was placed into a ground truth book and filed for later use.

DATA PROCESSING/ANALYSIS

The initial phase of data processing dealt with registering the ground

ORIGINAL PAGE IS
OF POOR QUALITY

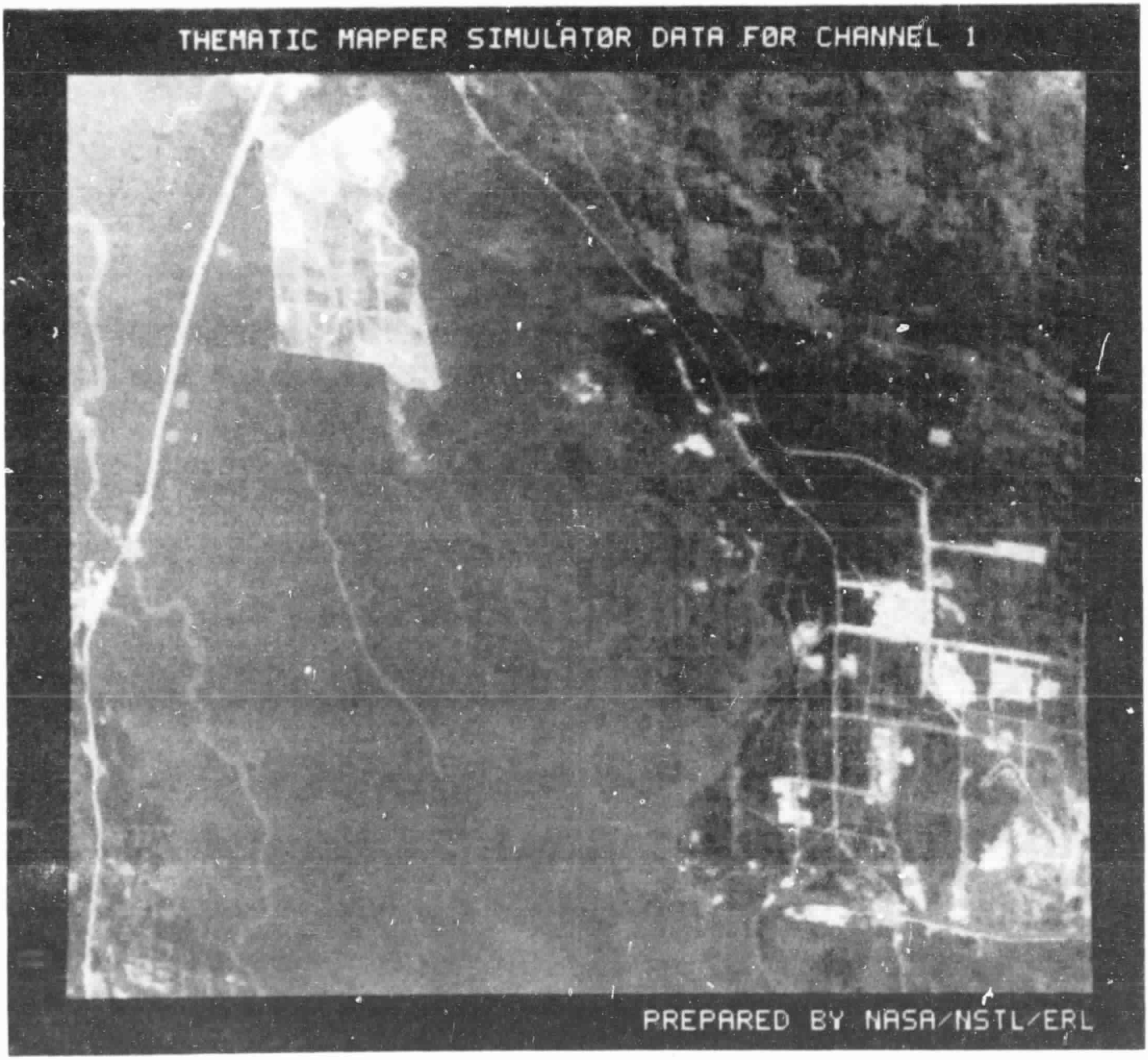


Figure 7. TMS Channel 1 Image

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

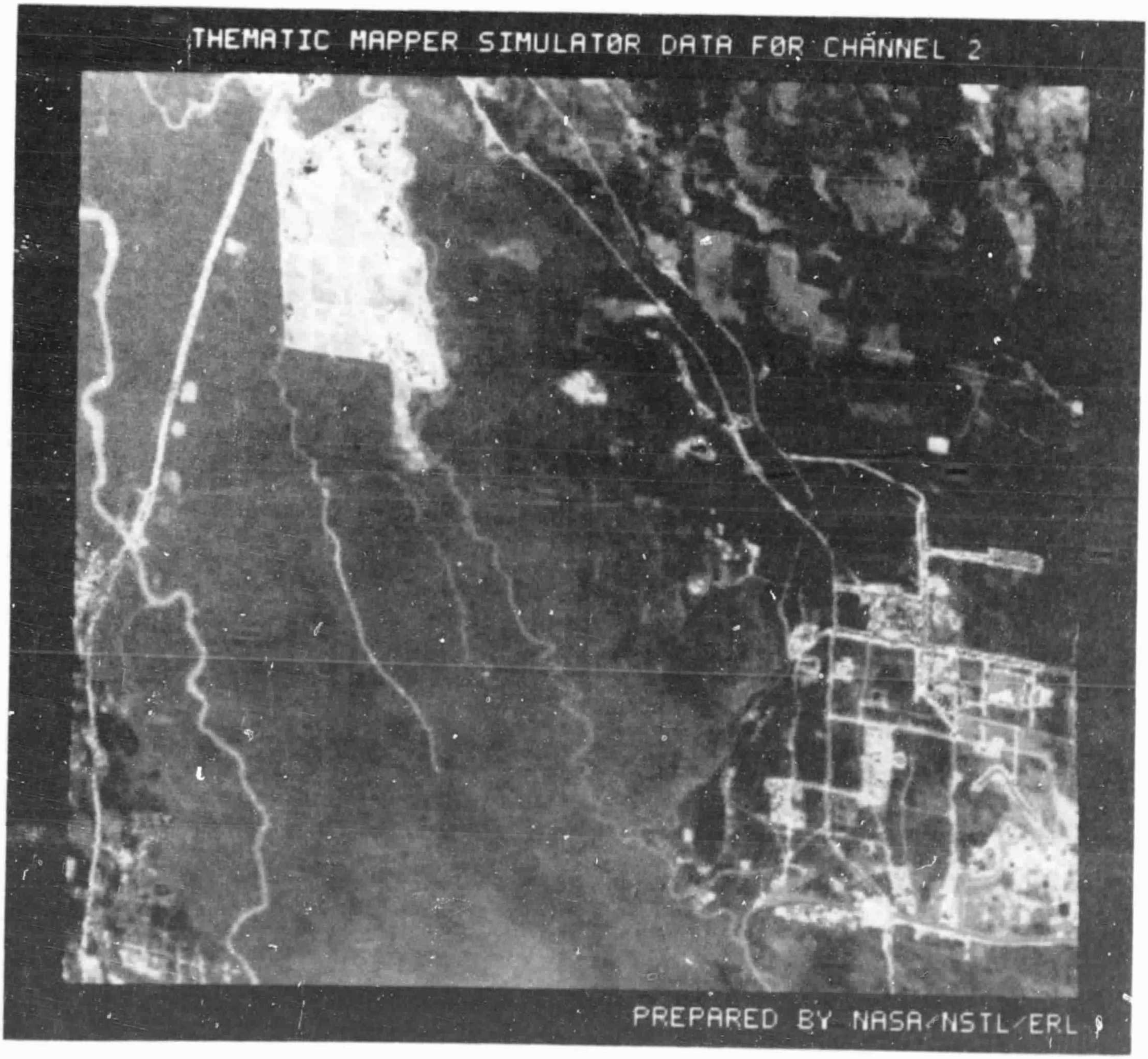


Figure 8. TMS Channel 2 Image

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

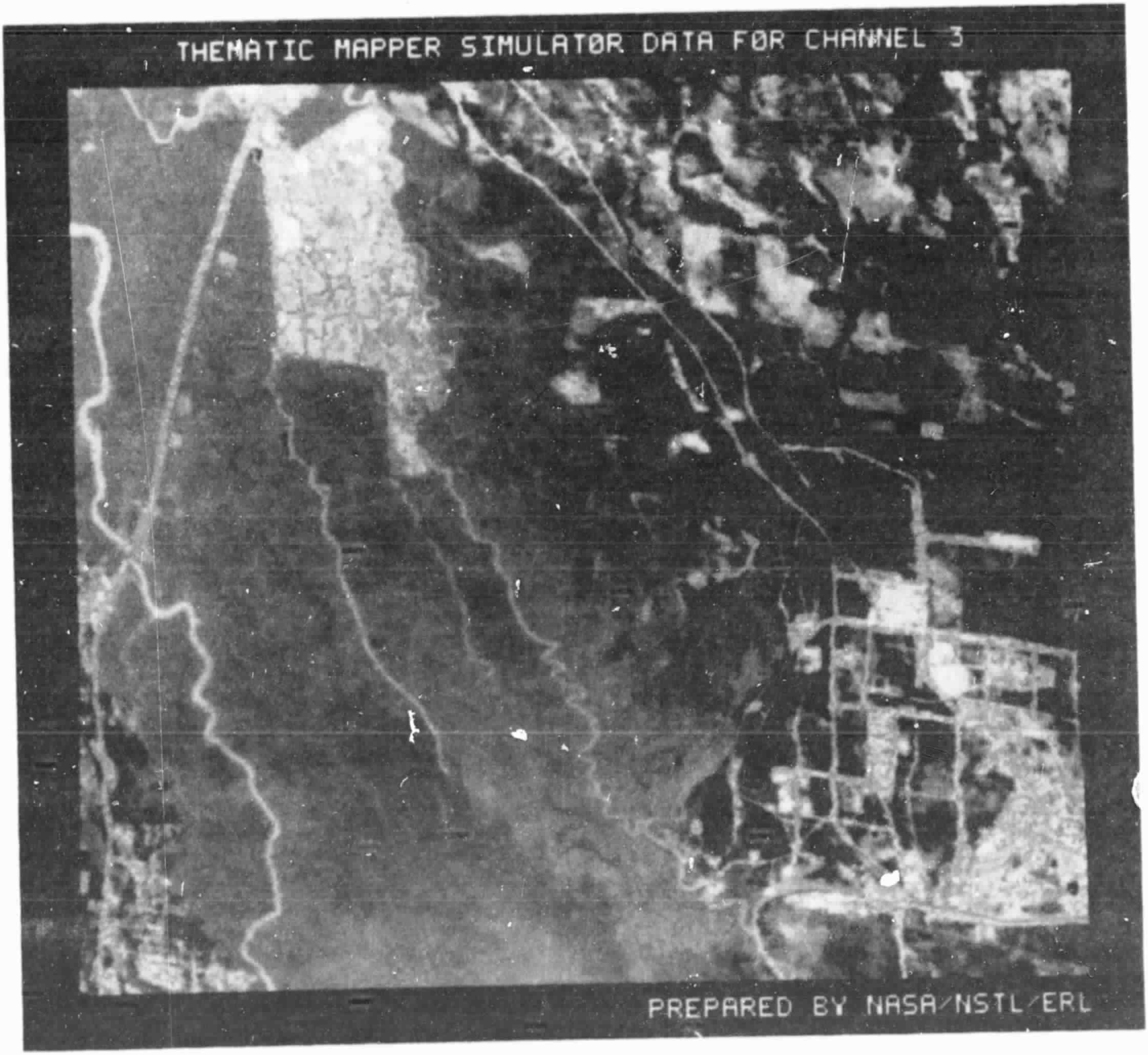


Figure 9. TMS Channel 3 Image

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

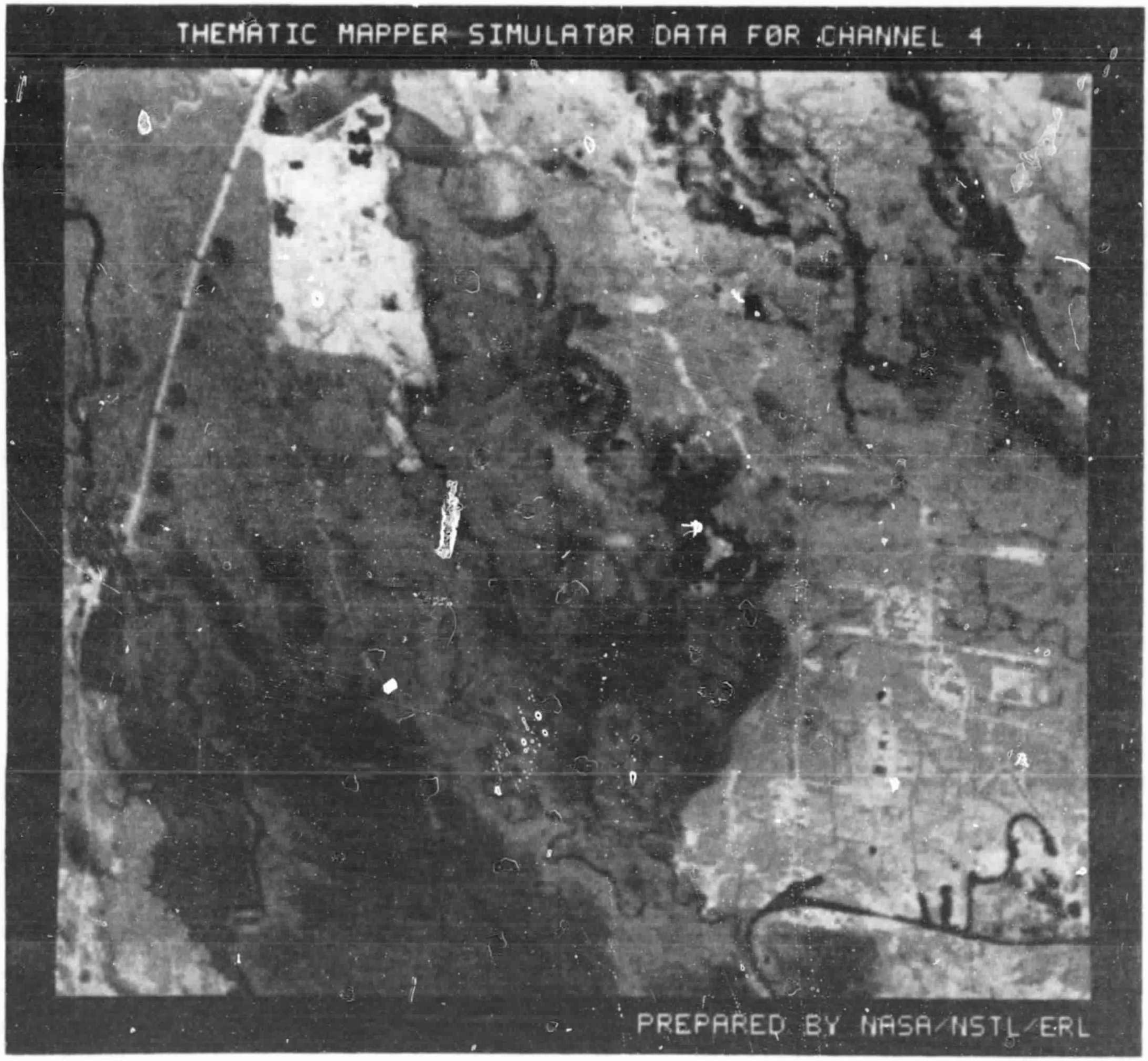


Figure 10. TMS Channel 4 Image

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

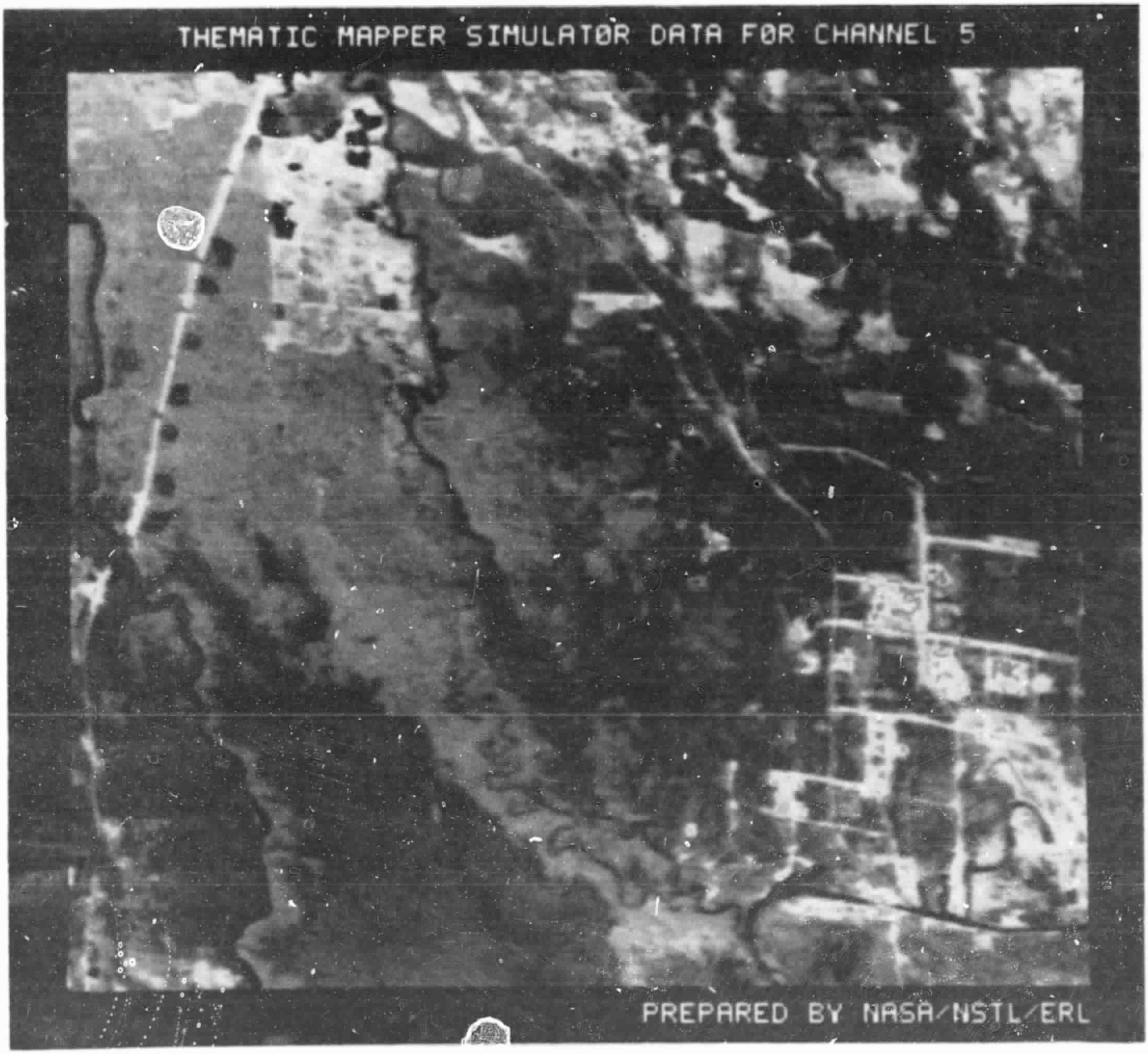


Figure 11. TMS Channel 5 Image

~~PRECEDING PAGE BLANK NOT FILMED~~

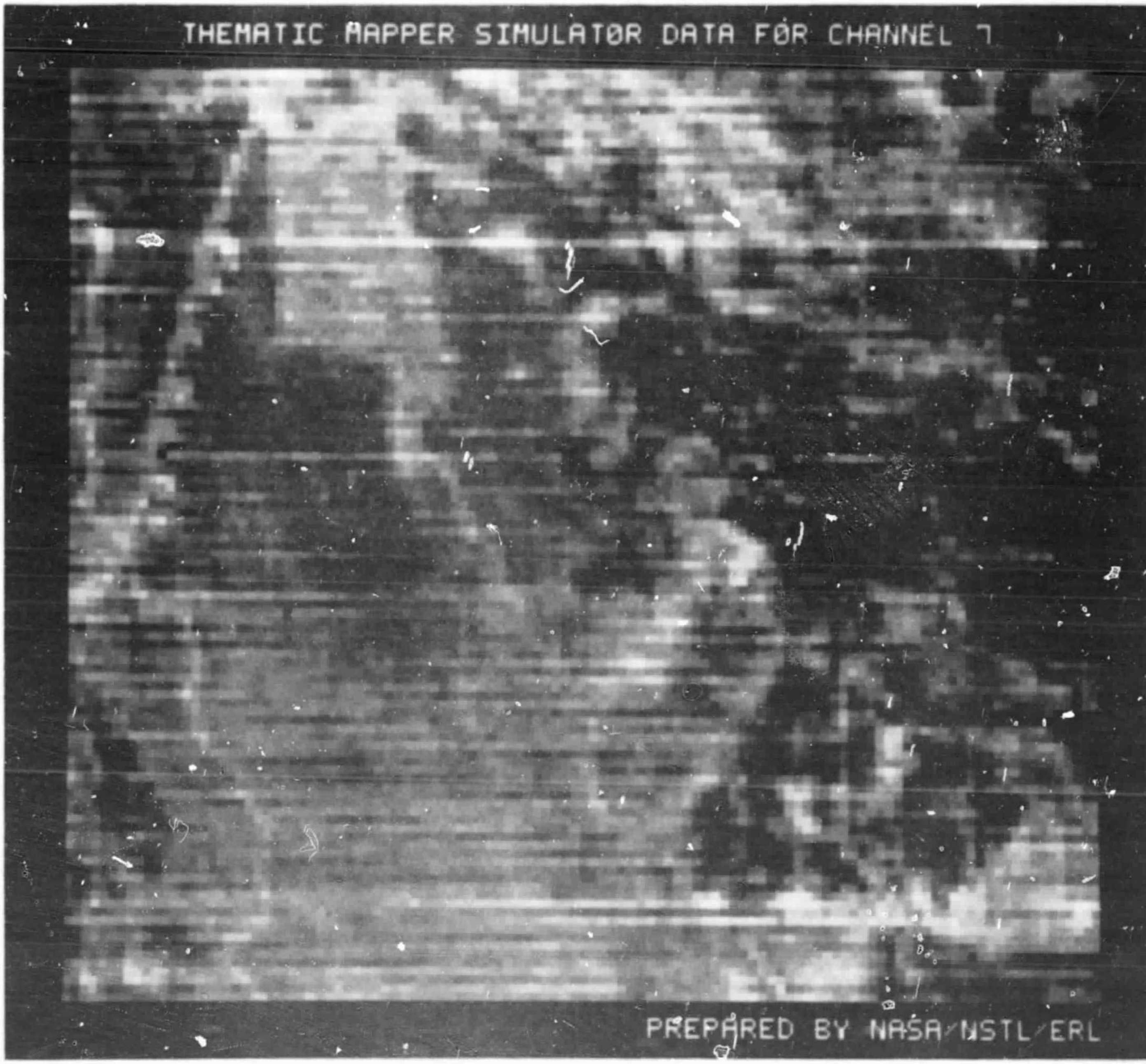


Figure 12. TMS Channel 7 Image

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

Table 1. Statistical Parameters Defining the Raw TMS Data Set Used in the Study

TMS CHANNEL	1	2	3	4	5	6	7
MEANS	37.81	92.87	85.43	79.93	67.72	46.22	105.21
STANDARD DEVIATION	6.53	22.05	29.96	28.17	29.54	12.45	14.49
COEFFICIENT OF VARIATION (X 100%)	17.27	23.74	35.07	35.24	43.62	26.94	13.77

truth data to the six channels of TMS data already located on a data file. The most straightforward manner in which to accomplish this was to create an additional channel of data for ground truth information. In this manner aerial photograph-to-UTM coordinate digitization and registration of the TMS data to a map base would not be necessary, thus eliminating air photo-to-map base registration errors.

Each ground truth site was located in the TMS data using aerial photographs collected concurrently with the digital data. Using an image display device, a polygon was outlined in the preprocessed TMS data which contained the TMS digital data corresponding to the site visited by field personnel. Polygon scan line/element corner coordinates corresponding to each ground truth site were stored for later use.

Upon completion of polygon selection, field sheets corresponding to each ground truth site/polygon pair were examined, and each polygon was given a unique identification value based on the major land cover type to which it was assigned. This value would serve as a cross reference between ground truth and TMS data, and help to make subsequent data analysis somewhat easier.

The "ground truth" channel of the data file containing the six-channel TMS data set was then initialized to contain zeros in all pixels. The coordinates for each polygon were recalled, one polygon at a time, and the value for all pixels in the ground truth channel within the boundary of the polygon being used was set to that of the corresponding polygon identification value. The processing of all polygons resulted in a channel of information containing solid polygons whose values correspond to those of the ground truth sites. Remaining pixels (background) retained the value of zero. By encoding the

ground truth channel in this fashion, results obtained could be analyzed on either a general land cover basis (by grouping) or, if desired, on a polygon-by-polygon basis.

The spectral homogeneity of all polygons was then examined for each of the six TMS channels. This was necessary in order to prevent misplacement of polygon corners from including dissimilar land cover types in the polygon. For example, several ground truth sites were located next to roads and adjacent to agricultural fields. Should the polygon be in error in such a manner that pixels from these land cover types were included in the polygon, evaluation of accuracy might lead to erroneous conclusions. Therefore, those pixels representing land cover types other than that intended for the polygon as a whole were edited out before the polygon was used.

In addition, since several polygons were located through the use of only one channel of TMS data, polygon editing furnished the opportunity to examine all remaining channels of preprocessed data. Certain problems unnoticed on the particular channel used, but discernible on one or more other channels, were corrected as a result.

Removal of the aberrant pixels (as determined through the use of polygon based histograms for each channel of data) was accomplished by simply zeroing them out in the ground truth channel of the combined data set, since all background (unused) pixels had been assigned the value of zero. Completion of this editing process resulted in spectrally pure ground truth for use in the analysis of TMS data performance.

SUPERVISED SPECTRAL SIGNATURE DEVELOPMENT

Having completed preprocessing of the TMS digital data, as well as registration (to the TMS data) and editing of ground truth polygons, the next step

in the investigation was to develop supervised spectral signatures for each of the edited ground truth polygons. Spectral signatures were developed through the use of software which uses a directional index table approach. This software can be instructed to examine one channel of data (a mask file) and to develop spectral values in the mask file. Thus, the ground truth channel was used as the mask file, and the software was instructed to develop a spectral signature for the edited ground truth polygons. Since every ground truth polygon in the ground truth channel had been assigned a unique value, individual spectral signatures were developed for each. This technique was required, since in some instances interior pixels of ground truth polygons had to be eliminated as described previously, while the original polygon boundary remained unchanged. Spectral signatures developed in this manner were stored in a disc file for subsequent use.

SUBSET SELECTION

The six-channel spectral signatures developed via the supervised approach were used to gain an initial understanding of the utility of each TMS channel. This was accomplished through the implementation of a subset selection technique. Average transformed divergence (reference 3), based on the supervised statistics developed from six channels of data, served as the vehicle for evaluation of overall performance of each subset of channels examined. As channels were removed, the resulting reductions in the average transformed divergence were noted. The subset of channels with the smallest reduction in average transformed divergence was selected as the most desirable one.

It is obvious that when using six channels of data, the number of subsets which must be examined would be quite large (68 in fact). In order to

circumvent this time consuming problem of examining 68 subsets, a technique using ordering of variables and properties of divergence was employed. This technique, detailed in reference 4, reduces the number of subsets which must be examined to a minimum. The results of subset selection are presented in Table 2.

Reduction in average transformed divergence can be used to gain some insight into the relative utility of the channel being deleted. For instance, the reduction in average transformed divergence for the five-channel subset resulting from the deletion of channel 1 was between 0.18 and 0.69 times that associated with the deletion of any other single channel. This indicates that the information content of channel 1, useful in improving the delineation of the land cover types in this study, is quite limited. Also, with such a small relative reduction, an investigator is more confident that this is the correct channel to delete to form the most useful five-channel subset. Had the reduction associated with the deletion of channel 1 been close to the reduction values for one or more of the other channels, more careful consideration would be necessary before the five-channel subset was formed.

This latter situation exists when considering the four-channel subset. In this instance, the reduction due to the elimination of channels 1 and 7 was 0.84 times that of the next closest subset of four channels. Clearly, removing channels 1 and 7 is, by definition, the proper course of action, but since the actual reduction in average transformed divergence is so close to the reduction associated with the next nearest subset, the decision is not without reasonable doubt. Thus, by computing the ratio of the reduction in average transformed divergence associated with the channels deleted to that of the next nearest subset of the same number of channels, a "confidence of decision" can be determined. High confidence implies that the

Table 2 Most Desirable Subsets of the Six Channels of Winter TMS Data from MS Study Site, Based on Average Transformed Divergence Values

NUMBER OF CHANNELS IN SUBSET	TMS CHANNELS	"CONFIDENCE" OF DECISION	PERCENT OF TOTAL REDUCTION
6 (original)	1,2,3,4,5,7	--	--
5	2,3,4,5,7	High	0.100
4	2,3,4,5	Moderate	0.410
3	2,4,5	Moderate	1.240
2	3,4	Low	3.880
1	4	Low	17.321

ratio is small (less than 0.70); moderate confidence that the ratio is between 0.71 and 0.80; and low confidence that the ratio is between 0.80 and 1.00. The confidence of decision values for the subsets selected are also presented in Table 2.

The desirability of creating subsets of various numbers of channels can be estimated by examining the percent of total transformed divergence lost by using the various subsets. This figure is computed as

$$\frac{\text{Average Transformed Divergence (6 channels)} - \text{Average Transformed Divergence (subset)}}{\text{Average Transformed Divergence (6 channels)}}$$

and is listed in the fourth column of Table 2. As expected, the percent reduction increases dramatically with the deletion of additional channels. The reduction associated with the five-channel subset is quite small, suggesting that results obtained with five channels would not be significantly different from those derived from six channels. The four-channel value is still small, but intuitively seems to be significant, and some doubt exists as to whether a four-channel subset should be created, since two four-channel subsets produced almost identical reductions in transformed divergence.

In order to test the level to which channels can be deleted without significantly reducing the utility of classification results, digital classifications were produced with a maximum-likelihood algorithm using the supervised spectral signatures corresponding to each subset listed in Table 2.

RESULTS

Accuracies (percent correct) for each classification produced were established based on an independent set of ground truthed polygons not included in spectral signature development.

The accuracy associated with water was, at this point in the analysis, found to be quite low (5.86 percent). All errors encountered were the result of water pixels remaining unclassified. Unclassified pixels were treated as errors in this study, since spectral refinement and editing had been performed on all ground truthed areas used in this investigation, including areas used for accuracy evaluation. This suggested that the turbidity-associated variance present with water was not adequately represented by ground truthed areas. This situation can be seen by careful examination of Figure 5.

Additional spectral signatures were established using channel 5 to define water/non-water, and were added to the spectral signatures previously developed. Channel 5 was used since spectral reflection in this region of the electromagnetic spectrum is very low, and all water, no matter how turbid, appeared very dark on a black and white display of the data. This action corrected the deficiency noted. Two-factor analysis of variance was then used to determine if significant differences (at the 95 percent level of confidence) existed between the overall results obtained from any of the subsets of channels used and those obtained from all six channels. The results of two-factor analysis of variance are presented in Table 3.

Column one of Table 3 presents the overall percent correct figures associated with each of the subsets listed in Table 2. These values were then subjected to the arcsin$\sqrt{P}$ transformation and the differences in transformed accuracies for each comparison listed in column two of Table 3 were computed (column three). Critical values were then established, and the differences were compared to them. Overall accuracies were determined to be significant if the difference in transformed accuracies between two subsets was greater than the critical value.

Table 3. Newman-Keuls Test Results of Two-Factor Analysis of Variance for Overall Results of Digital Data Classifications for Pearl River Basin TMS Data Set

Untransformed Accuracy (overall)	Subsets Compared	Difference in Transformed Accuracy	Critical Value	Significant 0.95 (*=yes)
6 ch 91.13	6 vs 5	0.792	1.521	
	6 vs 4	1.217	1.818	
	6 vs 3	5.451	1.521	*
	6 vs 2	9.627	1.818	*
	6 vs 1	34.369	1.993	*
5 ch 91.90	5 vs 4	0.425	1.521	
	5 vs 3	6.243	1.818	*
	5 vs 2	10.429	1.993	*
	5 vs 1	38.304	2.116	*
4 ch 92.30	4 vs 3	6.668	1.993	*
	4 vs 2	10.854	2.116	*
	4 vs 1	38.304	2.211	*
3 ch 85.01	3 vs 2	4.186	1.521	*
	3 vs 1	28.917	1.818	*
2 ch 87.06	2 vs 1	24.732	1.521	*
1 ch 38.30				

In this case, no significant difference was found to exist between overall results obtained from the four, five, and six-channel subsets. However, the one, two, and three-channel subsets were not only significantly different from the four, five, and six-channel subsets, but were significantly different from each other. This leads to the conclusion that based on overall results, no fewer than four channels of data should be used, as results are significantly compromised with three or fewer channels.

The results of the two-factor analysis of variance were also used to determine whether or not significant differences existed between classification results obtained for each of the land cover types for the six subsets of channels used. Results of this portion of the analysis of variance are summarized in Table 4.

The bars in Table 4 underline the best (highest percent correct classification) subsets for each land cover type which were found to be not significantly different at the 95 percent level of confidence. Thus, for any given land cover type, any of the subsets of channels underlined may be used with equally useful results. This, in effect, defines the reduction in dimensionality capable of being achieved for each land cover type listed.

It is of interest to note that no single subset of channels is underlined for all land cover types in Table 4. This implies that the results obtained for a given land cover type may be significantly compromised by the selection of channels, if such a selection were made on a "best overall" basis. Thus, based on Table 4, selecting all six channels would reduce the percent correct values for hay/grass and river bottom forest, while at the same time not significantly affecting the overall performance achieved for the classification as a whole.

Percent correct classification figures associated with each land cover type as well as for the classification as a whole are also presented in

Table 4. Results of Analysis of Variance Showing Statistically Non-Significant Subsets of TMS Data which Produced the Best Results for Each Land Cover Type (α -0.05, Winter Pearl River Basin Data Set). Subsets Connected by a Bar were Non-significant. Numbers Represent Percent Correct Values Associated with each Land Cover Type/Subset Combination.

TMS CHANNELS USED

LAND COVER	1,2,3,4,5,7	2,3,4,5,7	2,3,4,5	2,4,5	3,4	4
Inert	96.19	96.19	96.67	97.62	95.24	35.71
Hay/Grass	89.16	95.18	94.58	94.58	97.59	0
Old Fields	91.95	89.93	71.14	63.09	47.65	0
Marsh	89.29	89.29	89.29	75.00	64.29	0
River Bottom	86.41	91.29	95.54	76.98	87.73	63.79
Mixed Forest	92.54	90.15	85.37	82.69	80.34	55.22
Pine	89.29	89.03	91.58	88.52	71.30	16.58
Water	100.00	94.59	95.24	99.35	96.91	20.35
Overall	91.13	91.90	92.30	85.01	87.06	38.42

Table 4. For old fields, marsh, and mixed forest, the accuracies decrease with each successive elimination of a channel of information. This implies, where significant differences exist, that the particular channel deleted contained information which made a significant contribution to the delineation of the land cover type in question.

For the remaining land cover types, as well as the overall value, accuracies tend to increase initially, and finally drop off as additional channels are deleted. This apparently paradoxical situation resulted from pixels which were unclassified in the high dimensional cases (and hence tabulated as "errors") being correctly classified in reduced dimensions (thus increasing the percent correct values presented in Table 4.) No pixels were found which had been classified into other land cover types in the higher high dimensions and which, upon deletion of channels, were correctly classified in reduced data space. The conclusion here is that all of the variability of these land cover types was not represented in the supervised spectral signatures developed. Of course, this is a potential drawback to any supervised approach. In any event, reducing the dimensionality in cases such as these placed less restrictive limits on classification of pixels into land cover classes. A point is reached, however, beyond which classification performance is adversely affected with continuing deletion of channels, as is evidenced by Table 4.

Nonetheless, the reduced channel subsets performed acceptably well (in all cases at or above the 89 percent correct level) for at least one subset of channels. Specific results obtained begin to show significant degradation at various levels of channel reduction, depending on land cover; but when viewed over all land cover types, no fewer than four channels

should be used if concerned with the eight land cover types described in this report. Had river bottom forest behaved in a manner similar to mixed forest, the decision would be to use five channels as a minimum.

Recall that earlier in this report there was some reasonable doubt about forming the four-channel subset, based on a study of the reduction in average transformed divergence. In this case, the best four-channel subset had a ratio of 0.84 that of the next best subset of four channels (comparing the reductions in transformed divergence), which indicates that the utility of the best four-channel subset is only moderately better than the next best subset. Analysis of variance determined that in only one instance (river bottom forest) was a significant improvement realized by creating a four-channel subset, but that in two cases (old fields and mixed forest) a significant decrease in performance was noted. Other land cover types were not significantly affected. The choice, then, of creating a four-channel subset depends on the land cover type of interest; but based on all eight land cover types examined, the overall solution would be to use the five-channel subset. This decision reinforces the use of the relative reduction in transformed divergence as an indicator of the desirability of further channel reduction in the generation of subsets.

Of particular interest in this investigation is the performance of TMS data relative to the forest land cover types. In all three cases, percent correct classification was in the range of 85-95, depending on the subset selected. For pure stands of timber the four-channel subset (TMS channels 2, 3, 4, and 5) performed best for river bottom forest and was one of the non-significant best for pine (Table 4). However, when dealing with areas containing the attributes of both (i.e., mixed forest), at least one additional channel of information is needed in order to achieve the statistically best performance. Channel 7 was determined to be the next most useful

channel to add, and its addition to channels 2, 3, 4, and 5 produced a significant improvement in the percent correct value associated with mixed forest (Table 4). In this specific case, the addition of channel 7 reduced the confusion between mixed forest and pine, resulting in the correct classification of pixels which, in the four-channel subset, were classified as pine. Although the range of mean values for pine in channel seven (88.16 - 121.98) overlapped the range for mixed forest (88.64 - 96.39), the statistics of the pine classes which were confused with mixed forest in the four-channel subset were all located at the upper portion of the range of pine, and out of the range of mixed forest on channel 7. Thus, by adding channel 7, the confusion was significantly resolved.

ALTERNATE APPROACHES TO SPECTRAL SIGNATURE DEVELOPMENT

In addition to the supervised spectral signature development approach already mentioned. several unsupervised techniques were also examined, using the evaluation procedure already outlined for the supervised metnod. Fundamentally, unsupervised spectral signature development differs from supervised techniques in that unsupervised techniques "scan" the entire data set and, within limits established by the investigator, develop spectral signatures defining spectral cover types without prior knowledge of the land cover types which are contained within the data set. It then becomes a matter of relating the spectral signatures developed to land covers present, using aerial photographs, ground truth, etc. Once the spectral signature/land cover relationships have been established, evaluation of performance can be conducted.

Of the techniques for unsupervised spectral signature development to be found in the literature, two basic types were selected for inclusion in this investigation: the sliding window approach and point clustering. These two

were selected since they represent two basic approaches to unsupervised techniques, and because software (ELAS - reference 5) already existed for them on the computer system being used for this study. The NSTL/ERL ELAS computer program SRCH (SEARCH) is an unsupervised sliding window type approach to spectral signature development. A three by three pixel window is moved through the data set and, based on numerous parameters defining the operational limits of the software, spectral signatures are developed defining the spectral composition of the data used.

A modified version of SRCH, called SVCP (SRCH - variable channel parameters), permits the user to define spectral homogeneity independently for each channel of input data used. Since this would permit more reasonable assumptions to be made with respect to the relationships between the channels of input data, it was also included in the study.

Point clustering, another ELAS computer program (PTCL), employs techniques to develop spectral signatures by examining individual pixels of data. The frequency of sampling is input by the user. As each point is examined, a decision is made as to whether or not the new pixel is spectrally similar to points already examined. If it is, it is grouped with the similar pixel(s). If not, it remains as a separate spectral signature, and the next pixel in the data file is examined. The process continues until all data have been processed.

Various parameter settings of all of the unsupervised software were tried, and the results were compared to those obtained from the supervised approach discussed earlier. In all cases, the results obtained from the supervised approach were significantly better than the unsupervised approaches. This does not mean that unsupervised techniques will not perform as well as the supervised techniques for specific land cover types, but rather that,

when considered over the forest vegetation in this test site, the unsupervised techniques did not perform as well in their separation into the three classes employed. For instance, SRCH produced very high accuracy values for water, hay/grass, inert materials, river bottom forest, and pine, but did not develop any spectral signatures defining the mixed forest; thus, based on the objective of this investigation to deal specifically with the forest resource, SRCH failed to perform to the same level as the supervised approach.

It is of interest to note that no matter which of the techniques is used, ground truth polygons must be established to relate the spectral signatures to land cover types, and (using an independent set of polygons) to evaluate performance of the final results produced. Such areas must be spectrally homogeneous in order to prevent the introduction of error into the experiment. Since this is the case, the work required to incorporate ground truth into the data analysis framework is the same for supervised and unsupervised approaches.

CONCLUSIONS

As a result of the analysis of TMS data collected for the MS site in winter, the following conclusions can be made:

1. TMS data processed with supervised spectral signature development techniques can produce land cover classifications for inert, hay/grass, old fields, marsh, river bottom forest, mixed forest, pine forest, and water cover types at an overall accuracy of 92.3%, and for the three forest cover types of pine, mixed, and river bottom at accuracies of 91.5%, 92.5%, and 95.5% respectively.

2. Overall classification performance is affected by the number of channels used, and no fewer than four channels should be used (TMS channels 2, 3, 4, and 5).

3. Specific land cover results are affected by choice of channels used (subset feature reduction), and the choice of subset which maximizes the results obtained for one land cover type may adversely affect the results obtained for another land cover type.

4. The three forest types which predominate the MS site can be adequately delineated by use of supervised techniques. Unsupervised techniques, while producing results which were very good in other land cover types, could not delineate all three forest types with the same level of performance as the supervised technique.

It should be noted that the above conclusions relate to the eight general land cover types that were defined for this study. Additional research is being conducted in this test site to determine the capability to discriminate more detailed forest cover information (e.g., species, density, understory) with TMS data.

BIBLIOGRAPHY

1. USDA US Forest Service. 1963. Major Forest Types in the South. Southern Forest Experiment Station, New Orleans, LA.

2. White, D.A. 1979. Plant Communities of the Lower Pearl River Basin, Louisiana. Ph.D Dissertation, Tulane University, New Orleans, LA.

3. Swain, P.H. and S.M. Davis (eds). 1978. Remote Sensing: The Quantitative Approach. McGraw Hill Inc., Chapter 3.

4. Kalcic, M.T. 1978. Selection of the Best k-Subset of p-Variables in Discriminant Analysis, unpublished manuscript, Department of Experimental Statistics, Louisiana State University.

5. Graham, M.H. et. al., 1980. ELAS-Earth Resources Laboratory Applications Software. NASA/NSTL/ERL, NSTL Station, Mississippi 39529.